Fall 11-9-2019

Linking Library Resources to the Web through Linked Data

sankha subhra das
Manonmaniam Sundaranar University, India, ssdas91@gmail.com

Arpita Roy Chowdhury
Rabindra Bharati University, arprc015@gmail.com

Linking Library Resources to the Web through Linked Data

Arpita Roy Chowdhury
Research Scholar
Department of Library and Information Science
Rabindra Bharati University
Kolkata-700050, West Bengal
Mob: +91 8972983655, E-mail: arprc015@gmail.com
ORCID- https://orcid.org/0000-0002-9894-1903

Sankha Subhra Das
Research Scholar
Department of Library and information Science
Manonmaniam Sundaranar University
Tirunelveli-627012, Tamil Nadu
Mob: +91 7980000423, E-mail:ssdas91@gmail.com
ORCID- https://orcid.org/0000-0002-3480-6665

Abstract:

The concept of linked Data and Semantic Web have been gaining importance to revolutionise the world of the internet. The idea behind Linked Data can be used to create a huge database of the Internet interlinked by relationships that will be easily understood by machines as well as humans. We tried to compare some libraries that have published their content via linked open data. The advantages and disadvantages have also been highlighted. libraries should benefit by adopting linked Data but there are many challenges that will come up in its implementation. The BIBFRAME Initiative is a possible framework that will enable libraries to link their resources on to the web attracting users with high-quality content. Various initiatives have been undertaken by the libraries in order to make resources accessible to all users.

Keywords:

Linked data, Semantic web, database, Libraries, Internet, Interlink, BIBFRAME Initiative, framework

1. Introduction:

In the present digital era information is explored everywhere. Digital libraries are the moral centres for disseminating information. Growing digital knowledge expands the necessary information throughout the world. Where everyone can access information whenever and wherever they want. The main aim of digital knowledge is sharing and exploring the archived knowledge to users who are in immense requirement of it. Now-a-days online method of libraries such as data links, semantic content descriptors, and ontology is mashed up with social networking, wikis, social tagging etc. Applications have developed with enhancement of technology and metadata services. Linked data is such a type of service that can connect related data and make these information available on the web.

Here we have taken into account linked data services of various digital libraries discussing their advantage and disadvantage, using vocabulary etc.

2. Linked data:

Tim Berners Lee described linked data as a technique that will use standard web technologies to link related data on the web.

Tim Berners Lee, the inventor of the web, listed four principles for linked data:

1. Uniform Resource Identifiers (URIs) be used to identify things.

2. Use Hypertext Transfer Protocol (HTTP) Uniform Resource Identifiers (URIs) to look up things on the internet (interpreted, "dereferenced").

3. Utilizing open standards such as RDF and SPARQL, provide useful information about what a name identifies when it's looked up.

4. When publishing data on the web, refer to other things using their HTTP URI-based names.

However, in 2009 Berners-Lee offered three "extremely simple" rules for linked data:

1. All kinds of conceptual things, they have names now that start with HTTP.

2. If I take one of these HTTP names and I look it up . . . I will get back some data in a standard format, which is kind of useful data that somebody might like to know about that thing, about that event.

3. When I get back that information, it's not just got somebody's height and weight and when they were born, it's got relationships. . . . And when it has relationships, whenever it expresses a relationship, then the other thing that it's related to is given one of those names that starts with HTTP.

Data integration and linked data technology both express mentioned resources by a Uniform Resource Identifiers (URI) and express these relations using resource description and access (RDF). At present linked data has gained more importance as a topic in any digital library. Some of the models related to linked data have made a remark on it.

Tim Berners lee in 2010 proposed a 5star model to encourage people to publish more and more in a linked open data. The model is shown below:

Stars	**Focuses on**
1-star (*)	This type of data is available on the web, but with open license to be Open Data
2-star (**)	It is available as machine readable structured data
3-star (***)	It is available as non-proprietary format
4-star (****)	All the above facilities and use as open standards from world wide web consortium to identify things
5-star (*****)	All the above facilities including linking one’s data to others data providing context.

Tim Berner Lee’s 5-star Model (source Maria et al.:2015)

W3C's 5star data mug (source Coyle: 2012)

3. Related literature:

Das and others (2008) conducted a survey on Linked Data in scientific collaboration framework with popular content management system Drupal. They organised material in online communities by capturing semantic relationship.

Latif et al. (2010) made a case study on Open digital journal to discover and construct an author's profile from Linked data. They had developed a tool which can establish links between authors of digital journals with relevant semantic resources available on linked open data. They did a study on Journal of Universal Computer Science to examine their project. They found that their application can motivate researchers and developers also.

Chiarcos and others (2011) did a study on linked open data cloud. Open linguistics working group is an initiative from different fields. They experimented on these group and studied further.

Zaveri and others(2012) did a survey on quality assessment for linked open data. They presented the result of a systematic review for assessing quality of linked open data. A comparison has been made by them under a common classification list of the dimensions of metrics.

Neubert (2012) discussed Linked data based library web services in the field of economics building lightweight REST oriented web service interfaces mashed up with Web 2.0 application . He described how these services are integrated with publishing platform.

Coyle (2012) did an analysis on Semantic Web keeping in mind the concept of libraries data and common terminologies of the semantic web. He also discussed about the basic components of linked data, machine-readable data formats, open data and use of Uniform Resource Identifiers (URIs).

Konstantinou and others (2014) exposed bibliographic information as Linked Open data using Standards based mappings. They made their study on query capabilities and implementation of linked open data. Some of the databases also analysed for the purpose of data review.

Maria et al. (2015) made a study on linked data in digital libraries. Here they described about various national library and organization and their application of linked data in their digital libraries. The vocabulary which they have used are also discussed. Some of the benefits, problems and future are also mentioned.

4. Objectives:

1. To identify various digital libraries using linked data as their data sharable resource.
2. To find out the vocabulary used by these libraries.
3. To draw attention on the advantages of using linked data in digital libraries.
4. To highlight the disadvantages of using linked data in libraries.
5. To provide suggestions including the development of linked data.

5. Methodology:

The Data has been collected from various websites of libraries and databases like Google Scholar, J-store, SCOPUS, and DOAJ. The concerned websites were browsed thoroughly for extracting data according to the objectives defined in this study. The obtained data was analysed and reviewed properly.

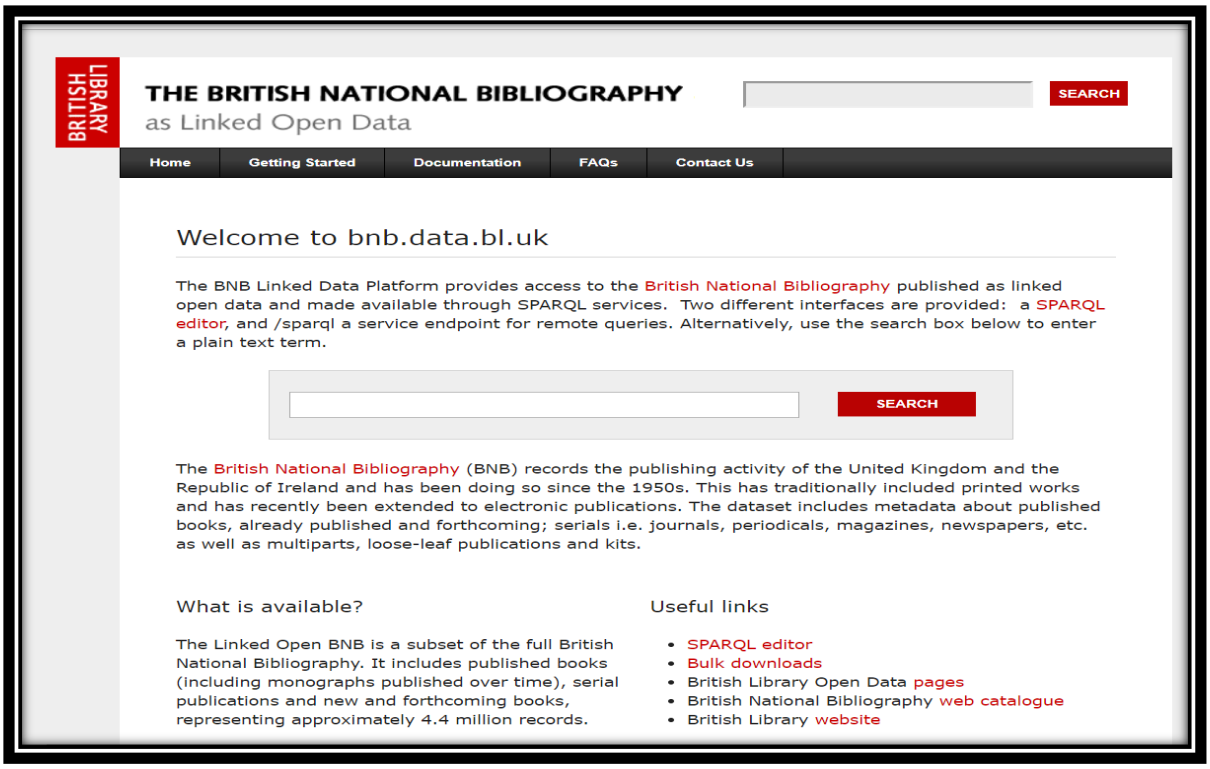
6. Some of the digital libraries connected via Linked data:

6.1. British library:

The British library is one of the oldest with a huge collection of documents. It has millions of items existing both in print and digital format. British library comprise a version of linked data. They use it for servicing others only for their books and serial publications. They created a model such as 'Things of interest' for reusing the most descriptive existing schemes.

The BNB Linked Data Platform allows one to access the British National Bibliography published as linked open data which can be availed through SPARQL services. Two different interfaces are provided: a SPARQL editor, and /sparql a service endpoint for remote queries. On the other hand, one can use the search box to enter a plain text term.

The Linked Open BNB is a subset of the full British National Bibliography. It includes published books and monographs published over time, serial publications, new and forthcoming books, representing an approximate of 4.4 million records.



Source: *Secondary data* (bnb.data.bl.uk)

The British National Bibliography dataset has been modelled and represented in RDF using a number of standard schemas. Their approach has been to re-use existing RDF vocabularies. Terms had been defined and they have been mentioned in the British Library Terms schema. Non-library domain specific schemas were also defined as much as possible.

Used Vocabularies	Data connected to other sources
BIBO(bibliographic ontology)	VIAF (Virtual International Authority File)
Bio(Biographic Information)	LCSH (Library of Congress subject headings declared in RDF)
Dublin Core (vocabulary of basic metadata terms)	LEXVO (standard language codes)
ISBD (international standard bibliographic description)	MARC,RDF Book mash up
Org (ontology organization)	Geo names (for country of publication)
RDF schema, OWL, FOAF, etc.	Dewey info

Source: *Secondary data* (bnb.data.bl.uk)

The following methods are available to search the British National Bibliography (BNB) published as Linked Open Data:

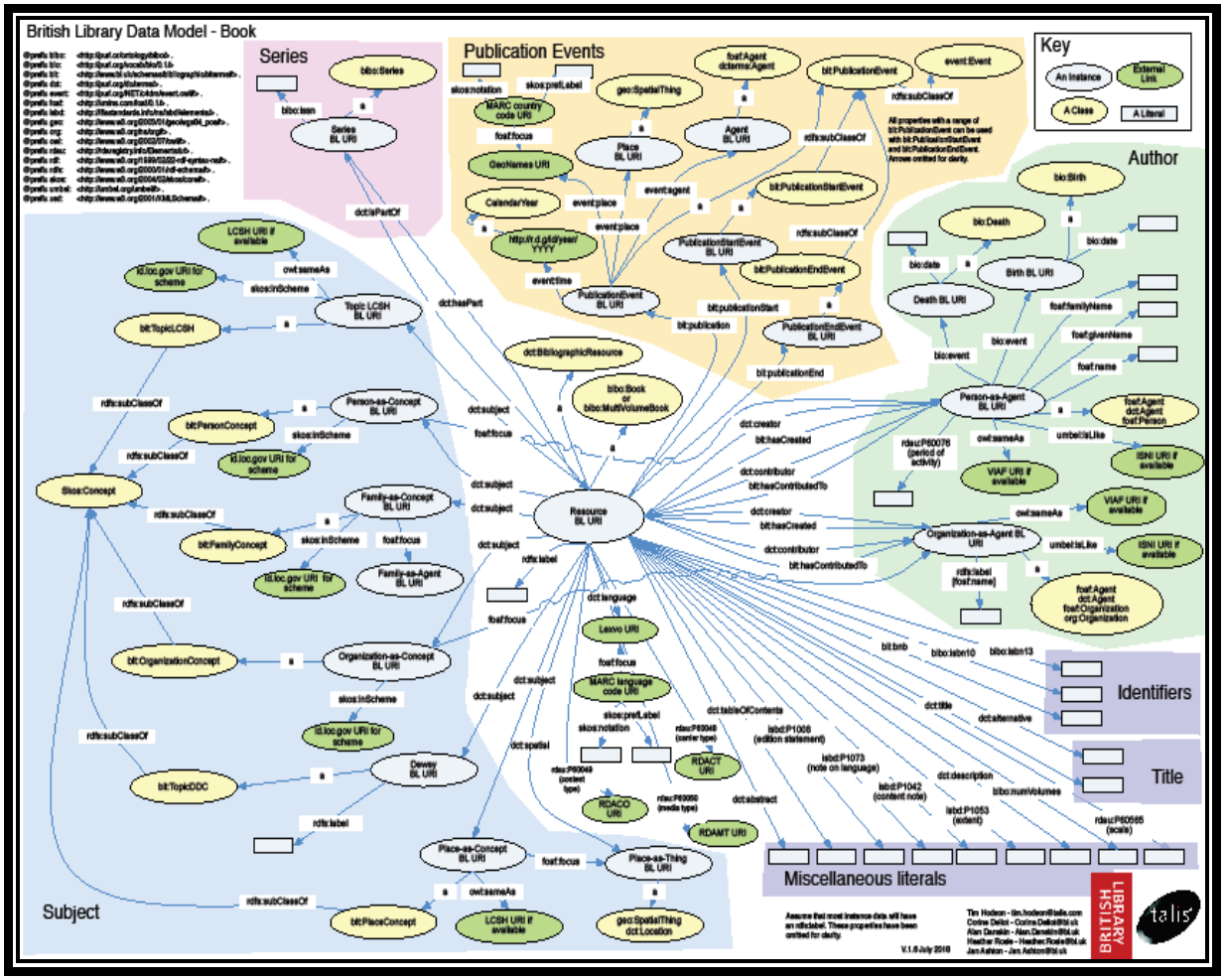
1. enter a plain text term in the search box,

2. use a SPARQL editor,

3. interrogate the SPARQL endpoint available at bnb.data.bl.uk/sparql.

SPARQL is a query language for RDF.

There are high-level overview diagrams of the data model, one for books mentioned below, one for serials and one for forthcoming publications.

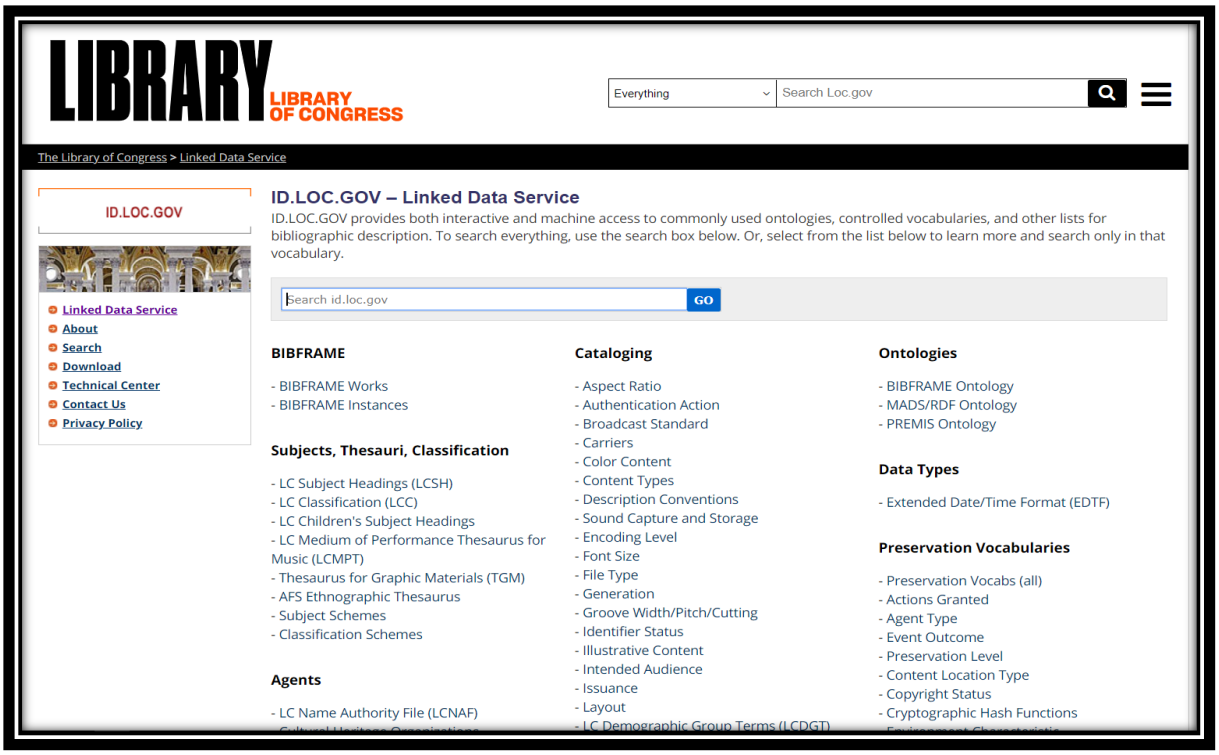


Source: *Secondary data* (http://www.bl.uk/bibliographic/pdfs/bldatamodelbook.pdf)

6.2. Library of congress:

The Linked Data Service (www.loc.gov) was by the Library of Congress in 2009. It has provided access to around 2,60,000 authority records. At present it also includes access to standards and vocabularies used by the Library of Congress. These are classification files, preservation vocabularies, themes, geographic areas, event types, countries, languages, , types of software, etc.

Since 2011, the Library of Congress implemented BIBFRAME model of Linked Data for integrating their data on the Web. The BIBFRAME model is further divide into classes such as authority, annotation, item and body. The relationship item reflects FRBR relationships in graphical terms rather than hierarchical relationships. The BIBFRAME model was defined in RDF, identifying all resources, attributes and relationships among properties, this allows mapping to other vocabularies. However there are certain connection between BIBFRAME and vocabularies such as SKOS, Dublin Core, FOAF.



source: *secondary data* (id.loc.gov)

id.loc.gov provides both interactive and machine access to commonly used ontologies, controlled vocabularies, and other lists for bibliographic description. A search option has been provided with a list below to search only in that vocabulary.

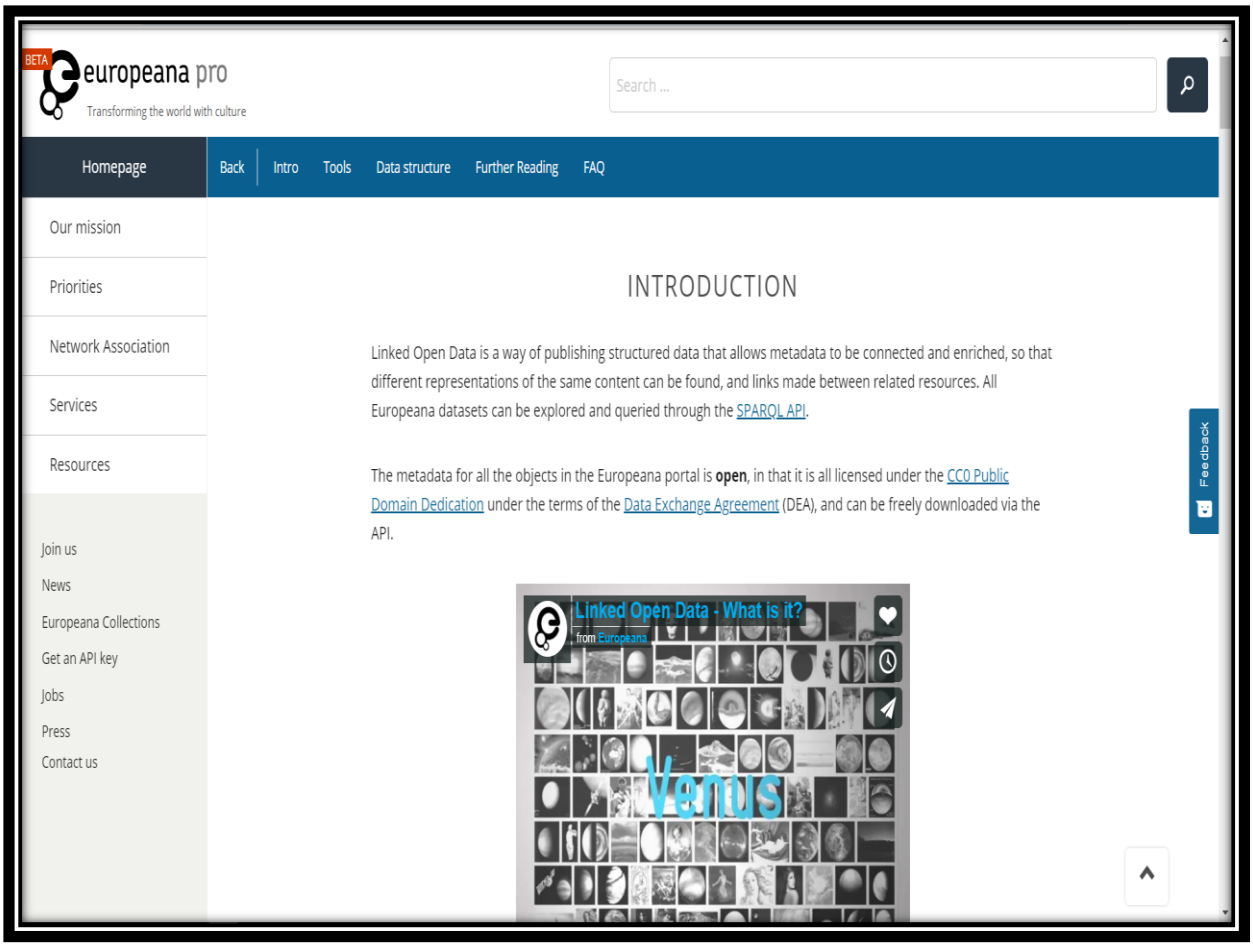
Used Vocabularies	Data connected to other sources
BIBFRAME (Bibliographic Framework)	VIAF (Virtual International Authority File)
FOAF (Friend-of-a-Friend)	LCSH (Library of Congress subject headings declared in RDF)
Dublin Core (vocabulary of basic metadata terms)	Extended Date/Time Format (EDTF)
PREMIS, MARC, MODS	MARC Geographic Areas (GAC)
METS, and MIX	JSON (MADS/RDF and SKOS/RDF)
RDF schema	SKOS Concept Scheme Offsite link

source: *secondary data* (id.loc.gov)

6.3. Europeana Pro:

Europeana (data.europeana.eu) is an European digital library that provides European institutions access to multilingual publications and linked metadata. This library includes millions of items digitized from libraries, archives, museums and multimedia collections.

The EDM (Europeana Data Model) is based on the principles of the Semantic Web and Linked Data. The model uses RDF standards and control vocabularies such as: OAI-ORE (Open Archives Object Reuse and Exchange), SKOS, Dublin Core, FOAF.



source: *secondary data* (https://pro.europeana.eu/page/linked-open-data)

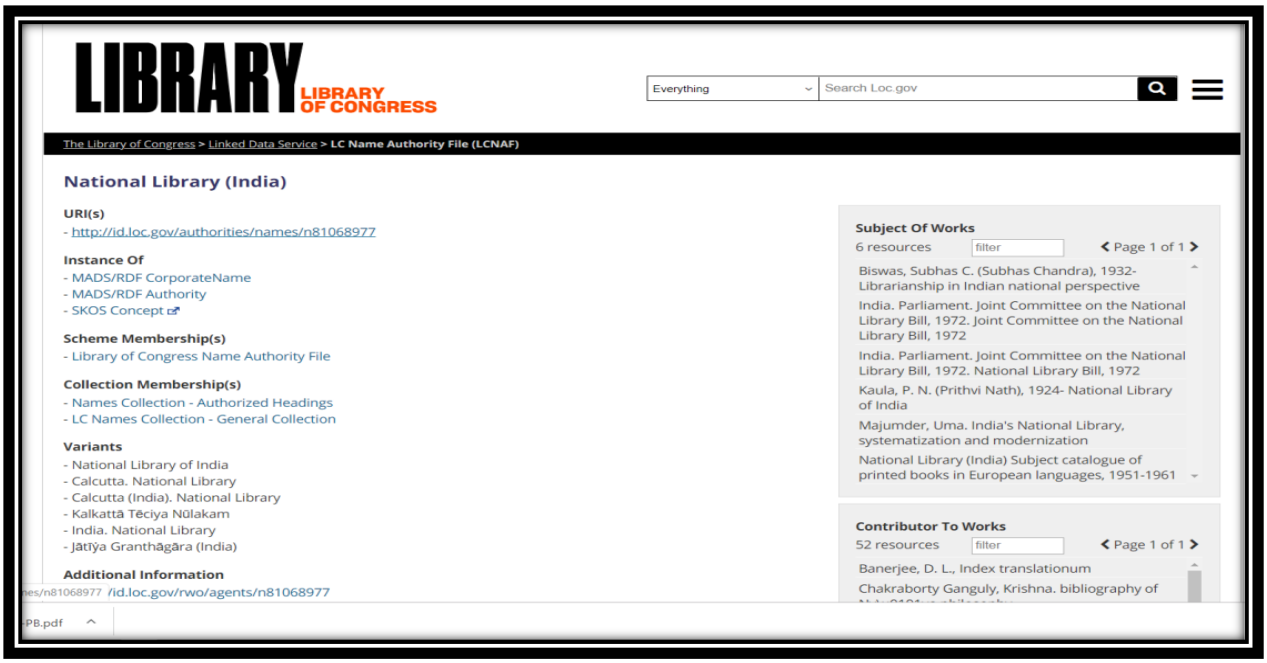
Used Vocabularies	Data connected to other sources
BIBFRAME (Bibliographic Framework)	VIAF (Virtual International Authority File)
FOAF (Friend-of-a-Friend)	
Dublin Core (vocabulary of basic metadata terms)	Extended Date/Time Format (EDTF)
RDF (Resource Description Framework) schema	MARC Geographic Areas (GAC)
	JSON (MADS/RDF and SKOS/RDF)
	SKOS (Simple Knowledge Organization System)Concept Scheme Offsite link

source: *secondary data* (https://pro.europeana.eu/page/linked-open-data)

EDM includes descriptive as well as contextual properties that includes features of resources and relate them to others. For descriptive metadata, EDM has builders that represent objects or events metadata. A pilot project has been undertaken that interlinks to several datasets like DBpedia, Geonames gazetteer, GEMET thesaurus. These concepts use SKOS.

6.4. National Library of India:

The National Library of India is associated with Library of Congress and provides Linked Open Data Service in collaboration with Library of Congress name authority file.



source: *secondary data* (http://id.loc.gov/authorities/names/n81068977.html)

Used Vocabularies	Data connected to other sources
BIBFRAME (Bibliographic Framework)	VIAF (Virtual International Authority File)
RDF (Resource description framework) schema	LCSH (Library of Congress subject headings declared in RDF)
Dublin Core (vocabulary of basic metadata terms)	MARC/XML
MARC (Machine Readable Cataloguing)	MARC Geographic Areas (GAC)
MADS/RDF Authority	JSON (MADS/RDF and SKOS/RDF)
MADS/RDF Corporate Name	SKOS Concept Scheme Offsite link

source: *secondary data* (http://id.loc.gov/authorities/names/n81068977.html)

7. Comparison of features related to Linked Data in the selected libraries.

Library	Benefits	Vocabularies and ontologies	Problems	Future
Europeana	It Provides queries to several linked metadata from European institutions.	RDF,OAI_ORE, SKOS, FOAF, Dublín Core	difficulty in migrating data to EDM.	Increasing involvement of the community with new datasets.
Library of Congress USA	Model people, places and events were related using existing schemes.	BIO DUBLÍN CORE, ISBD, ORG, SKOS, RDF SCHEMA, OWL, FOAF, RDA.	Inconsistency of experts in different areas for the transformations.	Browsers can be developed on BIBFRAME. Migration tools need to be developed.
British Library	Shape things of interest related to a bibliographic resource such as people, places,	BIBO, BIO, Dublin Core, ISBD, SKOS, RDF Schema, FOAF.	Lack of applications consuming Linked Data.	More Improvement on data linking tools need to be established.

	events, themes. Improved visibility.			

source: *secondary data*

7.1. Advantage:

- The visibility of the data can be improved.
- links can be established via online services. .
- The Open data recovery can also be improved.
- Various query on linked metadata can be consulted from multiple institutions.
- The end user resources annotations improve their credibility

7.2. Disadvantage:

- Numerous vocabularies are there for the same metadata,
- The difficulty of migrating data onto new models.
- An increasing need to develop tools for transforming to Linked Data.
- Shortage of experts in different areas related to transformations.
- There is still indicators lacking about usage of Linked Data.

7.3. Suggestions:

- Enhancement of community participation in cataloguing.
- The participation of the community should be must with new datasets and links to external sources.
- New model browsers and data migration tools need to be developed.
- The data links should to be defined.
- Better tools should be available for mapping data links.
- Feedback about data usage must be addressed.
- Existing library data standards must be compatible with Linked Data.

8. Conclusion:

Linked Data can be considered as a framework that works with Semantic Web, brought out by Tim Berners Lee, who gave the idea of turning the Internet into one large database rather than simply an unrelated compilation of documents. As the Internet occupies the most preferred option where users rely on information, hence libraries must take advantage of the concept behind Linked Data. Libraries can upload their resources onto the web, which can be visible to users, and this will in turn attract readers to the library for accessing authoritative and high quality resources.

9. References:

1. bnb.data.bl.uk (accessed October 16, 2019)

2. Chiarcos, C. and others(2011). Towards a linguistic linked open data cloud: The open linguistics working group.TAL.v.52(3),p.245-275

3. Coyle, K.(2012) Linked Data Tools: Connecting on the Web. Library Technology reports.

4. Das and others(2008).Linked data in a scientific collaboration framework. World wide web conf.08April,Beijing, China.

5. EDUCAUSE Review 54, no. 1 (Winter 2019). © 2019 Sayeed Choudhury. The text of this article is licensed under the Creative Commons Attribution 4.0 International License.

6. Gonzales, Brighid M. (2014). Linking Libraries to the Web: Linked Data and the Future of the Bibliographic Record. Information Technology And Libraries, p. 10-22.

7. http://www.bl.uk/bibliographic/pdfs/bldatamodelbook.pdf (accessed October 19, 2019)

http://id.loc.gov/authorities/names/n81068977.html (accessed October 21, 2019)

8. id.loc.gov (accessed October 21, 2019)

9. Konstantinou and others(2014).Exposing bibliographic information as linked open data using standards based mappings: methodology and results. Procedia-social and behavioural sciences.v.147,p.260-267

10. Landis, Cliff. (2014). A web of meaning: Linked open data resources on the web. College & Research Libraries News, v. 75(9), p. 492-505,ISSN 2150-6698. Available at: <https://crln.acrl.org/index.php/crlnews/article/view/9193/10174>. doi:https://doi.org/10.5860/crln.75.9.9193 (accessed October 12, 2019)

11. Latif, A.(2010)Discovery and Construction of Authors' Profile from Linked Data :A case study for Open Digital Journal.*LDOW2010*, April 27, 2010, Raleigh, USA.

12. Maria, H. and others(2015).Current state of Linked Data in Digital Libraries. http://jis.sagepub.com/content/early/2015/07/18/0165551515594729.abstract

13. Neubert, J.(2012).Linked data based library web services for economics. International conf. on Dublin Core and Metadata applications,2012.

14. "Transforming Library Metadata into Linked Library Data", American Library Association, September 25, 2012.http://www.ala.org/alcts/resources/org/cat/research/linkeddata)Document ID: fc4b530d-00bb-4aef-b6d4-21d40f4ae19c (accessed October 8, 2019)

15. Zaveri, A. and others(2012).Quality assessment for linked open data: A survey.Undefined1IOS press,p.1-5.